

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Bandman, Olga
Guegler, Karl J.
Corley, Neil C.
Shah, Purvi
- (ii) TITLE OF THE INVENTION: NEW HUMAN CATHEPSIN
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
 - (B) STREET: 3174 Porter Drive
 - (C) CITY: Palo Alto
 - (D) STATE: CA
 - (E) COUNTRY: USA
 - (F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: DOS
 - (D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: To Be Assigned
 - (B) FILING DATE: Herewith
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Billings, Lucy J.
 - (B) REGISTRATION NUMBER: 36,749
 - (C) REFERENCE/DOCKET NUMBER: PF-0331 US
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 415-855-0555
 - (B) TELEFAX: 415-845-4166
 - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 334 amino acids

PF-0331-3 DIV

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: THYMNOT02
- (B) CLONE: 347021

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met	Asn	Leu	Ser	Leu	Val	Leu	Ala	Ala	Phe	Cys	Leu	Gly	Ile	Ala	Ser
1				5					10					15	
Ala	Val	Pro	Lys	Phe	Asp	Gln	Asn	Leu	Asp	Thr	Lys	Trp	Tyr	Gln	Trp
			20					25					30		
Lys	Ala	Thr	His	Arg	Arg	Leu	Tyr	Gly	Ala	Asn	Glu	Gly	Gly	Trp	Arg
		35					40					45			
Arg	Ala	Val	Trp	Glu	Lys	Asn	Met	Lys	Met	Ile	Glu	Leu	His	Asn	Gly
	50					55					60				
Glu	Tyr	Ser	Gln	Gly	Lys	Leu	Gly	Phe	Thr	Met	Ala	Met	Asn	Ala	Phe
65					70					75					80
Gly	Asp	Met	Thr	Asn	Glu	Glu	Phe	Arg	Gln	Met	Met	Gly	Cys	Phe	Arg
				85					90					95	
Asn	Gln	Lys	Phe	Arg	Lys	Gly	Lys	Val	Phe	Arg	Glu	Pro	Leu	Phe	Leu
			100					105					110		
Asp	Leu	Pro	Lys	Ser	Val	Asp	Trp	Arg	Lys	Lys	Gly	Tyr	Val	Thr	Pro
	115						120					125			
Val	Lys	Asn	Gln	Lys	Gln	Cys	Gly	Ser	Cys	Trp	Ala	Phe	Ser	Ala	Thr
	130					135					140				
Gly	Ala	Leu	Glu	Gly	Gln	Met	Phe	Arg	Lys	Thr	Gly	Lys	Leu	Val	Ser
145					150					155					160
Leu	Ser	Glu	Gln	Asn	Leu	Val	Asp	Cys	Ser	Arg	Pro	Gln	Gly	Asn	Gln
			165					170					175		
Gly	Cys	Asn	Gly	Gly	Phe	Met	Ala	Arg	Ala	Phe	Gln	Tyr	Val	Lys	Glu
		180						185					190		
Asn	Gly	Gly	Leu	Asp	Ser	Glu	Glu	Ser	Tyr	Pro	Tyr	Val	Ala	Val	Asp
	195						200					205			
Glu	Ile	Cys	Lys	Tyr	Arg	Pro	Glu	Asn	Ser	Val	Ala	Asn	Asp	Thr	Gly
	210					215					220				
Phe	Thr	Met	Val	Ala	Pro	Gly	Lys	Glu	Lys	Ala	Leu	Met	Lys	Ala	Val
225				230						235					240
Ala	Thr	Val	Gly	Pro	Ile	Ser	Val	Ala	Met	Asp	Ala	Gly	His	Ser	Ser
				245					250					255	
Phe	Gln	Phe	Tyr	Lys	Ser	Gly	Ile	Tyr	Phe	Glu	Pro	Asp	Cys	Ser	Ser
		260					265						270		
Lys	Asn	Leu	Asp	His	Gly	Val	Leu	Val	Val	Gly	Tyr	Gly	Phe	Glu	Gly
	275						280					285			
Ala	Asn	Ser	Asn	Asn	Ser	Lys	Tyr	Trp	Leu	Val	Lys	Asn	Ser	Trp	Gly
	290					295					300				
Pro	Glu	Trp	Gly	Ser	Asn	Gly	Tyr	Val	Lys	Ile	Ala	Lys	Asp	Lys	Asn
305					310					315					320
Asn	His	Cys	Gly	Ile	Ala	Thr	Ala	Ala	Ser	Tyr	Pro	Asn	Val		
			325						330						

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1366 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: THYMNOT02
- (B) CLONE: 347021

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CTCAGAGGCT	TGTTTGCTGA	GGGTGCCTGC	GCAGCTGCGA	CGGCTGCTGG	TTTTGAAACA	60
TGAATCTTTC	GCTCGTCC TG	GCTGCCTTTT	GCTTGGAAT	AGCCTCCGCT	GTTCACAAAT	120
TTGACCAAAA	TTTGGATACA	AAGTGGTACC	AGTGGAAGGC	AACACACAGA	AGATTATATG	180
GCGCGAATGA	AGAAGGATGG	AGGAGAGCAG	TGTGGGAAAA	GAATATGAAA	ATGATTGAAC	240
TGCACAATGG	GGAATACAGC	CAAGGGAAC	TTGGCTTCAC	AATGGCCATG	AATGCTTTTG	300
GTGACATGAC	CAATGAAGAA	TTCAGGCAGA	TGATGGGTTG	CTTTCGAAAC	CAGAAATTCA	360
GGAAGGGGAA	AGTGTTCCTG	GAGCCTCTGT	TTCCTTGATCT	TCCCAAATCT	GTGGATTGGA	420
GAAAGAAAGG	CTACGTGACG	CCAGTGAAGA	ATCAGAAACA	GTGTGGTTCT	TGTTGGGCTT	480
TTAGTGCGAC	TGGTGCTCTT	GAAGGACAGA	TGTTCCGGAA	AACTGGGAAA	CTTGTCCTCAC	540
TGAGCGAGCA	GAATCTGGTG	GACTGTTCGC	GTCCTCAAGG	CAATCAGGGC	TGCAATGGTG	600
GCTTCATGGC	TAGGGCCTTC	CAGTATGTCA	AGGAGAACGG	AGGCCTGGAC	TCTGAGGAAT	660
CCTATCCATA	TGTAGCAGTG	GATGAAATCT	GTAAGTACAG	ACCTGAGAAT	TCTGTTGCTA	720
ATGACACTGG	CTTCACAATG	GTCGCACCTG	GAAAGGAGAA	GGCCCTGATG	AAAGCAGTCG	780
CAACTGTGGG	GCCCATCTCC	GTTGCTATGG	ATGCAGGCCA	TTCGTCCTTC	CAGTTCCTACA	840
AATCAGGCAT	TTATTTTGAA	CCAGACTGCA	GCAGCAAAAA	CCTGGATCAT	GGTGTCTTGG	900
TGGTTGGCTA	CGGCTTTGAA	GGAGCAAATT	CGAATAACAG	CAAGTATTGG	CTCGTCAAAA	960
ACAGCTGGGG	TCCAGAATGG	GGCTCGAATG	GCTATGTAAA	AATAGCCAAA	GACAAGAACA	1020
ACCACTGTGG	AATCGCCACA	GCAGCCAGCT	ACCCCAATGT	GTGAGCTGAT	GGATGGTGAG	1080
GAGGAAGGAC	TTAAGGACAG	CATGTCTGGG	GAAATTTTAT	CTTGAAACTG	ACCAAACGCT	1140
TATTGTGTAA	GATAAACCA	TTGAATCATG	GAGGATCCAA	GTTGAGATTT	TAATTCTGTG	1200
ACATTTTAC	AAGGGTAAAA	TGTTACCACT	ACTTTAATTA	TTGTTATACA	CAGCTTTATG	1260
ATATCAAAGA	CTCATGCTT	AATTCTAAGA	CTTTTGAAAT	TTCATTTTTT	AAAAAGATGT	1320
ACAAAACAGT	TTGAAATAAA	TTTAAATTCTG	TATATAAAAA	AAAAAA		1366

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 333 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: 29715
- (B) CLONE: GenBank

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met	Asn	Pro	Thr	Leu	Ile	Leu	Ala	Ala	Phe	Cys	Leu	Gly	Ile	Ala	Ser
1				5					10					15	
Ala	Thr	Leu	Thr	Phe	Asp	His	Ser	Leu	Glu	Ala	Gln	Trp	Thr	Lys	Trp
			20					25					30		

2

Lys	Ala	Met	His	Asn	Arg	Leu	Tyr	Gly	Met	Asn	Glu	Glu	Gly	Trp	Arg
		35					40					45			
Arg	Ala	Val	Trp	Glu	Lys	Asn	Met	Lys	Met	Ile	Glu	Leu	His	Asn	Gln
	50					55					60				
Glu	Tyr	Arg	Glu	Gly	Lys	His	Ser	Phe	Thr	Met	Ala	Met	Asn	Ala	Phe
65					70					75					80
Gly	Asp	Met	Thr	Ser	Glu	Glu	Phe	Arg	Gln	Val	Met	Asn	Gly	Phe	Gln
				85					90					95	
Asn	Arg	Lys	Pro	Arg	Lys	Gly	Lys	Val	Phe	Gln	Glu	Pro	Leu	Phe	Tyr
			100					105					110		
Glu	Ala	Pro	Arg	Ser	Val	Asp	Trp	Arg	Glu	Lys	Gly	Tyr	Val	Thr	Pro
		115					120					125			
Val	Lys	Asn	Gln	Gly	Gln	Cys	Gly	Ser	Cys	Trp	Ala	Phe	Ser	Ala	Thr
	130					135					140				
Gly	Ala	Leu	Glu	Gly	Gln	Met	Phe	Arg	Lys	Thr	Gly	Arg	Leu	Ile	Ser
145					150					155					160
Leu	Ser	Glu	Gln	Asn	Leu	Val	Asp	Cys	Ser	Gly	Pro	Gln	Gly	Asn	Glu
				165					170					175	
Gly	Cys	Asn	Gly	Gly	Leu	Met	Asp	Tyr	Ala	Phe	Gln	Tyr	Val	Gln	Asp
			180					185					190		

Asn	Gly	Gly	Leu	Asp	Ser	Glu	Glu	Ser	Tyr	Pro	Tyr	Glu	Ala	Thr	Glu
		195					200					205			
Glu	Ser	Cys	Lys	Tyr	Asn	Pro	Lys	Tyr	Ser	Val	Ala	Asn	Asp	Thr	Gly
		210				215					220				
Phe	Val	Asp	Ile	Pro	Lys	Gln	Glu	Lys	Ala	Leu	Met	Lys	Ala	Val	Ala
225					230					235					240
Thr	Val	Gly	Pro	Ile	Ser	Val	Ala	Ile	Asp	Ala	Gly	His	Glu	Ser	Phe
				245					250					255	
Leu	Phe	Tyr	Lys	Glu	Gly	Ile	Tyr	Phe	Glu	Pro	Asp	Cys	Ser	Ser	Glu
			260					265					270		
Asp	Met	Asp	His	Gly	Val	Leu	Val	Val	Gly	Tyr	Gly	Phe	Glu	Ser	Thr
		275					280					285			
Glu	Ser	Asp	Asn	Asn	Lys	Tyr	Trp	Leu	Val	Lys	Asn	Ser	Trp	Gly	Glu
		290				295					300				
Glu	Trp	Gly	Met	Gly	Gly	Tyr	Val	Lys	Met	Ala	Lys	Asp	Arg	Arg	Asn
305					310					315					320
His	Cys	Gly	Ile	Ala	Ser	Ala	Ala	Ser	Tyr	Pro	Thr	Val			
				325					330						

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 334 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: 1468964
- (B) CLONE: GenBank

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Lys	Pro	Ser	Leu	Phe	Leu	Thr	Ala	Leu	Cys	Leu	Gly	Ile	Ala	Ser
1				5					10					15	
Ala	Ala	Pro	Lys	Leu	Asp	Gln	Asn	Leu	Asp	Ala	Asp	Trp	Tyr	Lys	Trp
			20					25					30		
Lys	Ala	Thr	His	Gly	Arg	Leu	Tyr	Gly	Met	Asn	Glu	Glu	Gly	Trp	Arg
		35					40					45			
Arg	Ala	Val	Trp	Glu	Lys	Asn	Met	Lys	Met	Ile	Glu	Leu	His	Asn	Gln
		50				55					60				
Glu	Tyr	Ser	Gln	Gly	Lys	His	Gly	Phe	Ser	Met	Ala	Met	Asn	Ala	Phe
65				70					75					80	
Gly	Asp	Met	Thr	Asn	Glu	Glu	Phe	Arg	Gln	Val	Met	Asn	Gly	Phe	Gln
				85				90					95		
Asn	Gln	Lys	His	Lys	Lys	Gly	Lys	Val	Phe	His	Glu	Ser	Leu	Val	Leu
			100					105					110		
Glu	Val	Pro	Lys	Ser	Val	Asp	Trp	Arg	Glu	Lys	Gly	Tyr	Val	Thr	Ala
		115					120					125			
Val	Lys	Asn	Gln	Gly	Gln	Cys	Gly	Ser	Cys	Trp	Ala	Phe	Ser	Ala	Thr
		130				135				140					
Gly	Ala	Leu	Glu	Gly	Gln	Met	Phe	Arg	Lys	Thr	Gly	Lys	Leu	Val	Ser
145					150					155					160
Leu	Ser	Glu	Gln	Asn	Leu	Val	Asp	Cys	Ser	Arg	Pro	Gln	Gly	Asn	Gln

PF-0331-3 DIV

				165					170					175			
Gly	Cys	Asn	Gly	Gly	Leu	Met	Asp	Asn	Ala	Phe	Gln	Tyr	Val	Lys	Asp		
			180					185					190				
Asn	Gly	Gly	Leu	Asp	Thr	Glu	Glu	Ser	Tyr	Pro	Tyr	Leu	Gly	Arg	Glu		
		195					200					205					
Thr	Asn	Ser	Cys	Thr	Tyr	Lys	Pro	Glu	Cys	Ser	Ala	Ala	Asn	Asp	Thr		
	210					215					220						
Gly	Phe	Val	Asp	Ile	Pro	Gln	Arg	Glu	Lys	Ala	Leu	Met	Lys	Ala	Val		
225				230						235					240		
Ala	Thr	Val	Gly	Pro	Ile	Ser	Val	Ala	Ile	Asp	Ala	Gly	His	Ser	Ser		
			245					250					255				
Phe	Gln	Phe	Tyr	Lys	Ser	Gly	Ile	Tyr	Tyr	Asp	Pro	Asp	Cys	Ser	Ser		
		260						265					270				
Lys	Asp	Leu	Asp	His	Gly	Val	Leu	Val	Val	Gly	Tyr	Gly	Phe	Glu	Gly		
	275					280						285					
Thr	Asp	Ser	Asn	Ser	Ser	Lys	Phe	Trp	Ile	Val	Lys	Asn	Ser	Trp	Gly		
	290					295					300						
Pro	Glu	Trp	Gly	Trp	Asn	Gly	Tyr	Val	Lys	Met	Ala	Lys	Asp	Gln	Asn		
305				310						315					320		
Asn	His	Cys	Gly	Ile	Ser	Thr	Ala	Ala	Ser	Tyr	Pro	Thr	Val				
			325					330									